ISO/IEC JTC1/SC29/WG1 N90025
90th Meeting – Online – 18-22 January 2021

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION

ISO/IEC JTC 1/SC 29/WG1
(ITU-T SG16)

Coding of Still Pictures

JBIG	**JPEG**
Joint Bi-level Image Experts Group	Joint Photographic Experts Group

TITLE: JPEG Fake Media: Context, Use Cases and Requirements

SOURCE: JPEG (ISO/IEC JTC 1/SC 29/WG 1)

PROJECT: JPEG Fake Media Exploration

STATUS: Approved

REQUESTED ACTION: Distribution

DISTRIBUTION: Public

Contact:
ISO/IEC JTC 1/SC 29/WG1 Convener – Prof. Touradj Ebrahimi
EPFL/STI/IEL/GR-EB, Station 11, CH-1015 Lausanne, Switzerland
Tel: +41 21 693 2606, Fax: +41 21 693 7600, E-mail: Touradj.Ebrahimi@epfl.ch

JPEG Fake Media:
Context, Use Cases and Requirements

1 Executive Summary

Recent advances in media manipulation, particularly deep learning based approaches, can produce near realistic media content that is almost indistinguishable from authentic content to the human eye. These developments open opportunities for creative production of new content in the entertainment and art industry. However, they also lead to the risk of the spread of manipulated media such as 'deepfakes'. This may lead to copyright infringements, social unrest, spread of rumours for political gain or encouraging hate crimes.

Clear annotation of media manipulations is considered to be a crucial element in many usage scenarios. However, in malicious scenarios the intention is rather to hide the mere existence of such manipulations. This already triggered various governmental organizations to define new legislations and companies (in particular social media platforms or news outlets) to develop mechanisms that can detect and annotate manipulated media contents when they are shared.

The scope of JPEG Fake Media is the creation of a standard that can facilitate secure and reliable annotation of media asset generation and modifications. The standard shall support usage scenarios that are in good faith as well as those with malicious intent.

2 Context & Motivation

Recent advances in media manipulation, particularly deep learning based approaches, can produce near realistic media content that is almost indistinguishable from authentic content to the human eye. These developments open opportunities for production of new types of media contents that are useful for the entertainment industry and other business usage, e.g., creation of special effects or artificial natural scene production with actors in the studio. However, this also leads to issues relating to fake media generation defying the integrity of the media (e.g., deepfakes), copyright infringements and defamation to mention some. Misuse of manipulated media can cause social unrest, spread rumours for political gain or encourage hate crimes. In this context, the term 'Fake Media' is used here to refer to any generated or modified media, independently of its 'good' or 'bad' intention, as well as media used in a misleading way.

In many application domains, the 'media manipulators' may want to declare the type of manipulations performed, in opposition to other situations where the intention is to hide the mere existence of such manipulations. This is already leading various governmental organizations to plan new legislation. Companies (especially social media platforms or news outlets) are developing mechanisms that would clearly detect and annotate manipulated media when they are shared. While growing efforts are noticeable in developing technologies, there is a need to have a standard for the media content and associated metadata, e.g., a JPEG standard that facilitates a secure and reliable annotation of fake media, both in good faith and malicious usage scenarios. Therefore, it is important for the JPEG Committee to better understand the fake media ecosystem and needs in terms of standardization through an in-depth analysis of fake media use cases, naturally independently of the 'intentions'.

It is envisaged that JPEG initiates a standardization activity in order to ensure interoperability between a wide range of applications dealing with fake media. To reach this goal, and as a first step, stakeholders are invited to join this effort by helping to better understand applications and scenarios relevant to fake media use cases. This will allow the JPEG Committee to then identify key requirements for a standard in fake media. Initial findings suggest that a set of standard mechanisms to signal fake media along with relevant information on the latter are needed. In addition, standard mechanisms for security and protection of integrity of media assets are desired. The latter is closely related to issues highlighted in media blockchain under progress in the last two years in JPEG and therefore is considered as a natural continuation of that effort.

It is also important to understand, in more depth, the usage scenarios which will require input from relevant industries, public bodies (responsible for legislations), technology providers and end-users. Therefore, the JPEG Committee has the intention to engage with stakeholders in this use case in order to develop a clearly defined roadmap for standardization.

3 Scope

The scope of JPEG Fake Media is the creation of a standard that can facilitate secure and reliable annotation of media asset generation and modifications. The standard shall support usage scenarios that are in good faith as well as those with malicious intent.

4 Definitions

To ensure a correct understanding of the descriptions in this document, this section defines terms and concepts as they are used in the context of this work.

- Fake Media: any generated or modified media asset, independent of its ‘good’ or ‘bad’ intention, as well as media used in a misleading way.
- Media asset: digital assets including images, videos, audio or text. In the context of this document we mainly focus on images, however, other media types are not necessarily excluded from the scope.
    - Media asset content: the content of the media asset excluding metadata, for example the pixel data in case of images.
    - Media asset metadata: data associated with the media content, such as annotations or IPR information.
    - Media asset origin: the method or device that created the media asset.
- Misinformation: false or inaccurate information that is communicated regardless of an intention to deceive.
    - Disinformation: a species of misinformation that is deliberately deceptive.
- Modification: changes made to a media asset.
    - Manipulation: modification with the intention to induce misinterpretation.
- Generated media asset: artificially created media asset.
- Media asset integrity: internal consistency or lack of corruption of a media asset.
- Media asset authenticity: traceable provenance of a media asset.
- Original media asset: media asset that has not been modified since its origin.
- Media provenance: trail of modifications made to a media asset.

5 Use cases

The JPEG committee currently identified use cases related to the following topics:

- Misinformation and disinformation
    - Deepfakes
    - Manipulated media
    - Media intentionally used out of context
- Forgery / Media forensics
    - Document forgery (e.g. IDs and passports)
    - Insurance fraud (e.g. pictures of accidents)

    - KYC (Know Your Customer) (e.g. fake identity)
    - Impostoring (e.g. impersonating a celebrity)
- Media modification
    - Image editing software
    - Movie preservation
    - Film enhancement
    - Restoration of old movies
- Media creation
    - Use of deep fakes for special effects
    - Green screens, media processing and composition
    - GAN (Generative Adversarial Network) images
    - Short content bursts
    - UGC (User Generated Content) e.g. TikTok, Triller, Adobe Spark
    - Media tracing, e.g. provenance, content versioning, context
    - Picture and movies production

Based on these topics, the following sections provide a preliminary overview of illustrative use cases. Both the topics and use cases will be extended in the future based on feedback from stakeholders.

5.1 Misinformation and disinformation

5.1.1 Media usage in breaking news

In his coverage, a journalist wants to use images from a social media post depicting police violence during protests. The journalist has to make a fast decision but of course he wants to be sure the image in the post is genuine and taken at the mentioned place and time.

5.1.2 Deep fake detection

A news host wants to double check if a video he received of the president making questionable claims is genuine and not a deep fake.

5.1.3 Content authenticity checking

An investigative journalist wants to verify if an image depicting past atrocities is actually from that era and place.

5.1.4 Content usage tracing

A photographer wants to find out where and how some of the images from his portfolio have been used and check whether they are used in an genuine context

5.1.5 Academic research

An academic journal reviewer might want to know if an image used as evidence for a successful experiment hasn't been altered and is accurately used.

5.1.6 Photographic framing

A journalist received images of the Grand Place in Brussels in the aftermath of the terroristic attacks. Due to the specific framing, the images give a frightening impression of the situation. Therefore, the journalist wants to compare with other images taken at the same place and time but from different perspectives to better evaluate the actual situation.

5.2 Media modification

5.2.1 Image colorization and restoration

A developer has created an algorithm that uses deep learning to colorize grayscale images and enhances the image quality. The output images are labelled to allow consumers to identify that these images have been processed and may not accurately reflect original colours.

5.2.2 Photo editing

A photographer uses photo editing software (e.g. Photoshop) to edit model pictures for a magazine. The final images are labelled to indicate that they are post-processed. The labels allow to signal how “severe” the changes are to distinguish simple contrast and tone enhancements from changes where content has been added, removed, modified or manipulated.

5.3 Media creation

5.3.1 Movie special effects

A creative movie production company has created several shots for a movie that are computer generated but almost indistinguishable from real footage. The generated footage is labelled to allow consumers to identify that the content is computer generated. Since the final movie is a composition of generated and real footage, the entire movie can be labelled frame by frame.

5.3.2 Media transcoding

A photographer develops multiple versions of an image for different purposes. This includes the camera RAW image, rendered JPEG, moderately enhanced image and varying quality versions for web preview or print. During each transcoding step, authenticity and IPR information is retained from the parent version to the child version. In addition, authenticity information might be updated to describe modifications inherent to the transcoding process such as loss of quality when transcoding to a lossy format.

5.3.3 Chroma keying or silhouette extraction

Using chroma keying or silhouette extraction, a reporter can be virtually placed in a different location. Labelling the content allows media consumers to identify whether the shots were actually taken at the location or not.

5.4 Forgery/media forensics

5.4.1 Insurance fraud

In the context of insurance fraud, an insurer might want to check whether an image used as evidence has not been manipulated.

5.4.2 Mileage reporting photo

A car insurance company provides a discount program for the customer of limited annual mileage and demands the annual-reporting photo showing the mileage and the time displayed on the front panel of the customer's car. This insurance company might want to check whether the photo reported has not been manipulated.

5.4.3 Photo for cost charge

A series of before & after photos is frequently used for charging repair-costs in modern digital society. In this case, the integrity of a series of photos with the timing information from the origin to the final needs to be authenticated.

5.4.4 Evidence of Trial

A prosecutor wants to verify whether a movie recorded by Closed Circuit TV Security System was really taken at the location and the time.

5.4.5 Media sharing on social media

A media consumer (end user) wants to verify the credibility of a news article shared on a social media and he/she would like to trace who created, modified and published the image on it.

5.4.6 Credibility of AI training image data sets

Online auction service buys a set of training image data from a stock photo service and wants to check if each image was really taken by a camera instead of being created synthetically.

6 Requirements

Although this is still preliminary, so far requirements in two main categories have been identified: modification description and secure signalling of authenticity information. The sections below list identified requirements for each of these categories.

6.1 Modification Description

- The standard shall provide means to **describe** how, by who or when the content was generated and/or modified.
- The standard shall provide means to describe the **type of modification**, e.g., no modifications, transcoded, enhanced, restored, colorized, edited, composed, deep fake, ...
- The standard shall provide means to describe the **purpose of a modification**.
- The standard shall provide means to describe (algorithmically or human) the **probability of a modification**
- The standard shall provide means to describe the **region** where the media asset was modified.
- The standard shall provide means to **attach provenance information** to media assets.
- The standard shall provide means to **keep track of the history of media asset modifications**.
- The standard shall provide means to **compress embedded descriptions**.
- The standard shall provide means to **embed references** to externally hosted descriptions.

6.2 Secure linking of modification descriptions and media content

- The standard shall provide means to **restrict access to** media asset metadata.
- The standard shall provide means to **identify** if the media asset has been modified.
- The standard shall provide means to record and protect **IPR information and/or provenance information**.
- The standard shall provide means to **identify the origin** of the media asset.
- The standard shall provide means to **verify the authenticity** of the media asset.